Texas A&M University System National Laboratories Office and Los Alamos National Laboratory partner to make large data sets easier to handle

May 4, 2020

LOS ALAMOS, N.M., May 5, 2020—The Texas A&M University System National Laboratories Office (NLO) and Los Alamos National Laboratory have formed a collaborative research effort to make extremely large data sets indexable and more easily searchable.

"We are excited to be partnering with our colleagues at Texas A&M on this important and potentially game changing research. This collaboration leverages extreme strengths in data management research from both our organizations," said Gary Grider, division leader for High Performance Computing at Los Alamos.

Fine-grain annotation of scientific data sets composed of trillions of mesh cells or particles is critical in making extremely large data sets indexable and more easily searchable. Typical multi-dimensional indexes analyze only the most interesting queries and typically require multiple passes over extremely large data sets and multiple days of dedicated processing time.

"We expect this collaboration to lead to the development of novel storage systems that will address high performance computing needs at the national labs," said Narasimha Reddy, the J.W.Runyon Professor in Electrical and Computer Engineering at Texas A&M University.

An efficient and resilient key-value interface with hardware support to accelerate fundamental operations offers significant promise to make the enormous data sets, generated by large-scale, multi-physics simulations as well as machine learning model training and inferencing, available for efficient analysis and provide scientists with powerful tools for both analyzing and understanding extreme-scale simulation data sets. Further, it protects data for long-term storage and enables efficient queries for data analysis at scale.

"Fine-grained annotation of scientific data sets is critical to enabling scientists to extract insight and understanding from the massive simulations we perform at Los Alamos. The key-value research we are engaged in with Texas A&M is a major element in unlocking that insight and accelerating the pace of scientific discovery," said Brad Settlemyer, senior scientist in the High-Performance Computing Design Group at Los Alamos.

This collaboration focuses on exploiting potential upcoming key-value flash devices to enable much more rapid analysis and insights of large data sets. It is based on recent work by Texas A&M in the area of hardware-assisted erasure protection for key-value stores. It is an excellent demonstration of utilization of processing power near/on storage devices to provide additional functionality and more powerful interfaces to applications.

This collaboration is jointly sponsored by the Efficient Mission Centric Computing Consortium (EMC3) and the NLO. In this consortium, High Performance Computing (HPC) consumer organizations, researchers and system developers can collaborate together to attack this challenging problem of higher efficiency extreme scale, mission-centric computing. The HPC consumer base, along with national and international HPC component and system developers, are encouraged to join EMC3.

For news and additional Information on LANL's EMC3 effort, visit usrc.lanl.gov/emc3.php

About the Texas A&M University System National Laboratories Office

The Texas A&M University System National Laboratories Office (NLO) was formed by the Chancellor to be a conduit for expanding engagement with the national laboratories for faculty, staff, and students of the A&M System. This office engages with all Department of Energy and National Nuclear Security Agency laboratories and sites.